**PSG COLLEGE OF TECHNOLOGY**

**Department of Applied Mathematics and Computational Sciences**

**IX Semester M.Sc DS 2020 - 2021**

**15XD96 Information Retrieval Lab**

**Package Abstract**

**Phase 1**

**Title:** Resume Short-listing

**Roll No:** 16PD22, 16PD36

**Name:** Nehal M, Subashini S

**Problem Statement:** Many institutions and companies use resumes as a way of short-listing candidates in preliminary rounds. Going through every resume and short-listing is an arduous process and in many instances, it is prone to human errors. This project aims to automate this short-listing process.

**Dataset:** Resumes and Job descriptions are collected from Naukri.com and linkedIn. Also job reviews are scraped from ambitionbox.com.

**Tool:** BeautifulSoup in python is used for scrapping documents from websites. NLTk and Spacy libraries are use to process documents.

**Implementation Details:**

* Data collected from resumes and job sections of Naukri.com and linkedIn.
* Data Cleaning
* Dictionary and Posting list are created.
* Then each resume is vectorized.
* Building search engine.
* Query is Job description given by the company
* Similarity score is computed between job description and resumes.
* Finally, search engine will rank and display the suitable candidates.
* Scope for improvement: incorporating NLP techniques like NER into the above model and improving on ranking methods.